Summer 6-5-2011

Making Robert Kane's Libertarianism More Plausible: How James Woodward's Interventionist Causal Theory can Give an Agent Control Over Her Undetermined Decisions

Tracy Van Wagner
Georgia State University

MAKING ROBERT KANE'S LIBERTARIANISM MORE PLAUSIBLE: HOW JAMES WOODWARD'S INTERVENTIONIST CAUSAL THEORY CAN GIVE AN AGENT CONTROL OVER HER UNDETERMINED DECISIONS.

by

TRACY P. VAN WAGNER

Under the direction of Dr. Eddy Nahmias

ABSTRACT

Robert Kane asserts that some decisions and actions which are made by an agent are undetermined. These undetermined decisions are what allow an agent to have free will and ultimate responsibility for her decisions and actions. Kane appeals to probabilistic causation in order to argue that these undetermined decisions are not arbitrary or random. I argue that Woodward's *interventionist* approach to causation can be used by Kane to make his theory of free will more plausible by illustrating how the agent causes her decision. Woodward's account can link an agent's reasons with her decision, activity in her self-network with her decision, and can render undetermined decisions plural rational, plural voluntary, and plural voluntarily controlled.

INDEX WORDS: Robert Kane, James Woodward, Libertarianism, Free will, Interventionist, Manipulationist, Intervention, Intervene, Control, Plural rationality, Plural voluntary, Plural voluntary control

MAKING ROBERT KANE'S LIBERTARIANISM MORE PLAUSIBLE: HOW JAMES WOODWARD'S INTERVENTIONIST CAUSAL THEORY CAN GIVE AN AGENT CONTROL OVER HER UNDETERMINED DECISIONS.

by

TRACY P. VAN WAGNER

A thesis submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2011

MAKING ROBERT KANE'S LIBERTARIANISM MORE PLAUSIBLE: HOW JAMES WOODWARD'S INTERVENTIONIST CAUSAL THEORY CAN GIVE AN AGENT CONTROL OVER HER UNDETERMINED DECISIONS.

by

TRACY P. VAN WAGNER

Committee Chair: Dr. Eddy Nahmias

Committee: Dr. Andrew J. Cohen

Dr. Dan Weiskopf

Electronic version approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

August 2011

DEDICATION

I would like to dedicate this thesis to my wonderful boyfriend, Howard, who put up with seeing very little of me and my high levels of stress throughout this project. Without his constant support and encouragement, I may not have made it through this project!

ACKNOWLEDGEMENTS

First, I would like to thank my committee chair, Dr. Eddy Nahmias, for all the hard work which he put into this project. I would also like to thank my committee members, Dr. Andrew J. Cohen and Dr. Dan Weiskopf for their helpful contributions. I would next to thank Dr. George Graham for providing very helpful insight and guidance on this project (as well as throughout my graduate student career). I would finally like to thank all of my friends and family for providing me with encouragement and support needed to keep me going.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter I

Introduction

A. Libertarian Free Will

In this paper, I will argue that James Woodward's *manipulationist* or *interventionist* account of causation should be used by Robert Kane in order to make his account of libertarian free will more plausible. I will be focusing on Kane's account of non-deterministic causation of willings and actions. In this section, I will give a brief introduction to libertarian free will by contrasting it with two other dominant views on free will: hard determinism and compatibilism.

Hard determinists argue that every action done by an agent is determined by antecedent events, the socialization, and the biological make up of the agent; these past conditions are sufficient to bring about the current action. In other words, it is not "up to us" whether we do action *A* or action *B*. So long as the determining conditions for action *A* come about, then the agent will do action *A*. This poses a problem for libertarian free will because, for libertarians, free will is not compatible with determinism and free will exists; it is "up to us" whether we do action *A* or action *B* for libertarians. (Kane 2005, 5-6)

Compatibilists argue free will is compatible with determinism. They argue for a different notion of freedom than hard determinists or libertarians. They argue that freedom is "a power or ability to do something." (Kane 2005, 13) and a lack of constraints to perform the actions we will to do and a lack of compulsions which interfere with our ability to will as we chose. Compatibilists argue that this sort of freedom is compatible with a determined past. We are free so long as nothing is restraining us from doing otherwise, should we will to do otherwise. (Kane 2005, 13-16)

Libertarians argue that free will is incompatible with determinism and that free will

exists. It is not the case, according to libertarians, that all of our actions and willings are determined by antecedent events. An action can occur or not occur "given *exactly the same past and laws of nature."* (Kane 1998, 106) A common criticism of libertarianism is that it is impossible or incoherent. If an event is not determined by antecedent events, then it seems that that event occurred randomly or by chance and is therefore not in anyone's control. A person having free will of this sort would be spastic and not in control of herself. In fact, the person would not have free "will" at all because she would not be in control of, or willing, the actions that she performs. One of Kane's goals is to address this argument, which he labels "The Intelligibility Question."

> The first step in doing this is to question the intuitive connection in people's minds between "indeterminism's being involved in something" and "its happening merely as a matter of chance or luck." *Chance* and *luck* are terms of ordinary language that carry the connotation of "its being out of my control." So using them already begs certain questions, whereas "indeterminism" is a technical term that merely precludes *deterministic* causation, not causation altogether. Indeterminism is consistent with nondeterministic forms of causation, where outcomes are caused, but not inevitably. It is a mistake to assume that *undetermined* means "uncaused." (Kane 2002, 421)

Kane focuses on indeterminate decisions made by agents. By appealing to neural networks, cognitive science, and psychology, Kane argues that agents can be in control of these indeterminate decisions. In brief, an indeterminate decision occurs when an agent holds two or more incommensurable reason sets each of which favor a different decision. The agent will deliberate and will eventually favor one reason set over the other(s). (Kane 1998, 16-17, Kane 2005, 33-35) In this paper, I will focus upon Kane's account of an agent's control over her indeterminate willings. Kane asserts that his theory needs an account of probabilistic causation, but he does not lay out an account which will give him what he needs for his theory to work: probabilistic causation by the self-network at the neural level of explanation, probabilistic

causation by the agent's mental states at the folk psychological level of explanation, decisions which satisfy the plurality conditions (plural rationality and plural voluntariness), and decisions which are plural voluntarily controlled. My goal is to find a theory of causation that will get Kane what he needs to make his libertarian view of free will more plausible.[1]

B. James Woodward's Interventionist Account of Causation

I propose that Woodward's *manipulationist* or *interventionist* approach to causation can provide the support needed to make Kane's theory more intelligible. Woodward argues that causal explanations consist of giving counterfactuals which contain information about how the initial conditions of a proposed causal claim, *X*, can be intervened upon in order to alter the possible effect, *Y*. If the proposed intervention is consistently correlated with a change in the proposed effect, then there is a causal relation between *X* and *Y*. These causal claims rely upon invariant generalizations, which are counterfactual generalizations about how a change in *X* will consistently cause a change in *Y.* (Woodward 2003, 9-15) I will propose that the laws of folk psychology are invariant generalizations. I will then argue that the indeterminate decisions which Kane requires in his theory arise when two or more of these generalizations which are embedded in a person's network of folk generalizations come into conflict. The input to each of these invariant generalizations is a set of reasons for performing each of the conflicting actions which the person wants to perform. I will focus upon Kane's example of a businesswoman on her way to an important meeting. She must get to this meeting on time in order to make a sale which is important to her career. However, she encounters a mugging in an alley on her way. If she stops to help, then she will be late for her meeting and will lose the sale. (Kane 1998, 126, Kane 2002, 417) Her reasons for going to the meeting are the input for her generalization about

[1] This paper is narrowly-focused on making Kane's theory of libertarian free will more plausible. I am not arguing that Kane's theory is the most plausible theory of free will, all things considered. That is a much larger debate and outside the scope of this paper.

advancing her career. Her reasons for stopping to help the victim are the input for her moral generalization. In Chapter IV, Section B, I will first illustrate how intervening on each of these reasons and generalizations shows how each chosen action relies upon these reasons and generalizations, thus giving an account of mental causation. Second, I will illustrate how interventions on the agents beliefs and desires shows that it is the simultaneous activation of these beliefs and desires that causes the conflict and deliberation, which shows that the deliberation is caused by the agent's set of beliefs and desires. Finally, using the conjunction of the businesswoman's beliefs and desires with their probabilistic outcome, I will give an account of the probabilistic causation of the decision. I will argue that these interventions show how the businesswoman's ultimate decision fulfills the plurality conditions and show the agent's plural voluntary control over her decisions. In Section C, I will run through similar arguments from Section B, by proposing similar interventions and drawing similar conclusions about the causation of the agent's decision by her self-network. In brief, I will argue that a decision is made when a network in the self-network receives sufficient electrical and chemical input. When two conflicting networks simultaneously receive chemical and electrical input, they conflict with each other, causing that area of the brain to move away from thermodynamic equilibrium. The two networks compete. The network which "wins" receives the needed electrical and chemical input to activate and the other network does not receive the sufficient chemical and electrical input to activate. I will run through the same interventions I ran through in Section B using generalizations involving neural networks and the chemical and electrical input as the input into those generalizations. In Section D, I will argue that Woodward's theory can operate on both the neural and the folk psychological level, thus giving Kane agency on the folk psychological level. He needs agency on the folk psychological level in order to make his

account an intelligible account of free will. In Section E, I will then argue that Woodward's theory can also give an account of deterministic non-SFAs.

C. Consideration of Other Theories

I will next consider two other theories of causation: David Lewis' counterfactual theory and Phil Dowe's conserved quantity theory. I will argue that Lewis' theory can give Kane what he requires, but in a more complicated fashion than Woodward's. I thus favor Woodward's account due to its simpler causal explanations for Kane. I will next argue that Dowe's theory cannot give a non-reductive account of causation at the folk psychological level. Using Dowe's account, the probabilistic causation of SFWs would have to propagate up from the neural level. This would mean that all indeterminacy occurs at the neural level and would take away some of the control of the person making the decision. The SFW would seem to be caused by the neural activity rather than the indeterminate deliberation of the agent. Woodward's account, in allowing indeterminacy on both the neural and psychological level, may diverge from Kane's original intent. Kane seems to rely upon the indeterminacy at the neural level propagating up to the psychological level in order to get indeterminate decisions at the folk level. (Kane 1998, 128-130) Using Woodward's account, Kane would be able to speak of indeterministic decisions at both the psychological and the neural level. This would seem to give the agent more control over the decision since, at the psychological level, the agent *qua* agent would be originating the indeterminacy of her decision rather than her neural networks. The agent would also be in control of the indeterminacy throughout her deliberation process, since she is in control of the deliberation process. Keeping the indeterminacy at the lower levels seems like it could easily fall prey to the above control criticisms: that the decision was made by her neural networks and she merely experienced the conflict without actively participating in the deliberation.

Chapter II

Kane's Libertarian Theory and What it Demands of Causation

A. Introduction

Free will, for Kane, "is *the power of agents to be the ultimate creators (or originators) and sustainers of their own ends or purposes."* (Kane 1998, 4) He is focused on freedom of the *will*, and refers to his position as a "free willist" view. In order for an agent to be the ultimate creator of her ends, her actions must be traced back to or originate from her own willing. Kane's view is a naturalistic libertarian view of free will. His view is naturalistic because he only uses categories or kinds that could also be used by compatibilist and hard deterministic theories of free will. A person's actions, for Kane, can be caused by her beliefs, motives, desires, and other mental states, which, according to Kane, exist in the natural order. He does not appeal to non-natural entities such as noumenal selves or non-natural forms of causation such as agent-causation. He argues that appealing to such non-natural entities and forms of causation that many libertarians make is one of the reasons that non-libertarians argue that libertarianism is unintelligible: it seems that we cannot give an account of an agent's ability to control her indeterminate decisions without appealing to entities and forms of causation outside of the natural order. (Kane 1998, 115-117, 136, 193, Kane 2002, 416). His view is libertarian because it requires that *some* of our choices be undetermined. Kane agrees that many of our freedoms are freedoms in the compatibilist sense in that our will is determined by antecedent events and our actions are free so long as they are not coerced or compelled. However, Kane argues that there are some freedoms which are incompatible with determinism. (Kane 1998, 15, Kane 2002, 408). The choices which are determined by one's character and motives must trace back to an undetermined choice (a choice for which one's character and motives are not a sufficient cause)

in order for an agent to possess free will. Kane refers to these undetermined actions as self-forming actions (SFAs) and the undetermined willings which cause the SFAs as self-forming willings (SFWs). (Kane, The Significance of Free Will 1998, 75, 125, Kane 2002, 408)

SFAs are actions performed by an agent which are not fully determined by her character, motives, upbringing, and surrounding circumstances. These SFAs help to form the agent's character; subsequent determined actions will be determined by the character and motives formed by the outcome of an agent's previous SFAs. For an SFA, the agent's will is not set before she acts. She has a choice between two or more competing sets of reasons or motives for each possible action. When an agent decides which set of reasons to favor, she is setting her will in favor of the option to which that set of reasons corresponds. The agent must make a mental effort to set her will. Kane refers to these acts of will as self-forming willings, or SFWs. SFWs "include such things as choices, decisions, judgments, formations of intention, and efforts or tryings, all of which are mental 'acts' or actions of one sort or another." (Kane 1998, 125, 75-78, Kane 2002, 407-437) (Kane 2002, 407-437)

For non-SFWs, the agent has sufficient motive to perform a certain action; this action is determined by the agent's past SFWs, current situation, and upbringing. The agent's will is already set by his character and motives; he does not have to make an effort to set his will one way or the other. Suppose Kane has decided to get a coffee from the coffee machine which has options for regular coffee and coffee with cream. Keeping with his usual habits, he has decided to get a regular coffee; his will is set one way for regular coffee. If he had deliberated about whether to get coffee or coffee with cream, the result, in this case, would have been Kane's discovering what he had already favored. It would not, like the SFWs, be a case of Kane's making the reason set for coffee prevail over the reason set for coffee with cream, because Kane

already had an all-things-considered preference for regular coffee. (Kane 2002, 412, 420, Kane 2005, 108-110)

B. Indeterministic Activity in the Neural Networks

Kane appeals to quantum theory and neuroscience in order to introduce genuine indeterminacy into his account. The "indeterminate efforts of will...are complex chaotic processes in the brain, involving neural networks that are globally sensitive to quantum indeterminacies at the neuronal level." (Kane 1998, 130) Conflicts in the agent's will cause the brain to move farther from thermodynamic equilibrium, which in turn amplifies the chaotic effects of quantum indeterminacies at the micro level. This increased sensitivity to changes at the micro level increases the influence these indeterminate processes have on the agent's decision. These processes are experienced by the agent as efforts of will and feelings of indecision. (Kane 1998, 130-131, Kane 2002, 422, 427) For non-SFWs, when our character and motives determine the decision, the indeterminacy does not arise in such a way that it can have an effect at the macro level. If it did, then it would be a fluke and would interfere in the agent's making rational decisions. (Kane 1998, 130-131, Kane 2002, 417, 422)

The reasons and intentions which are part of an agent's deliberation process are encoded in the motivational structure of the brain, or what Kane calls the "self-network." The self-network is a recurrent brain network which encodes the agent's plans and aspirations and guides the agent's practical reasoning. Kane proposes that the self-network is a set of neural networks distributed throughout the brain. This network is unified when the diverse groups of neurons cooperate to form "synchronized and causally interacting oscillations or wave patterns." (Kane 1998, 140) The self-network is activated for both indeterminate and determined decisions. (Kane 1998, 128-130, 137-142, 151, Kane 2002, 424)

When deliberating, an agent is considering two incommensurable sets of reasons: "It is as if the agent…is *trying*…to complete two competing (deliberative) tasks at once…Each task is being thwarted by the indeterminism coming from the other." (Kane 2002, 420) During deliberation, the agent does not have a preference for any of the sets of reasons. Her preference is only set once she makes her decision, which is when she favors one set of reasons over the other(s). (Kane 1998, 198) The conflicting reason sets for undetermined decisions correspond to two different neural networks within the self-network. Consider a businesswoman who witnesses an assault in an alley on her way to an important business meeting. When she pauses to deliberate, one of those neural networks encodes her desire to help the victim and her moral reasons for helping the victim. The other neural network encodes her selfish desire to get to the meeting and her reasons for wanting to get to the meeting. The indeterminacy of the decision results from the conflict between these two networks, which creates the tension experienced by the agent. Both of the networks are interfering with one another, each attempting to reach activation threshold at the same time. The result of the deliberation process will be one network reaching its activation threshold. This is when the agent makes her choice. (Kane 2002, 419, 422) The self-network is altered with each SFW. The reasons and intentions encoded in the self-network belong to the agent and can be given more or less weight (increased weight = increased chances of being favored in the future) depending upon the results of SFWs. If the businesswoman favors her moral reason set and helps the mugging victim, then that set of reasons will be given more weight in the future and she will be more likely to help someone in need in the future. In this way, the agent controls the formation and evolution of her self-network. (Kane 1998, 193-194)

In brief, it is the quantum indeterminism in the brain which makes SFWs indeterminate.

The reason sets that conflict indeterministically during deliberation are encoded in the self-network. It is this conflict which is experienced by the agent as her effort of will. The phenomenological experience of the agent is important for Kane's account. For Kane, the folk psychological explanation of an agent's deliberation is just as important as the neural explanation of her deliberation. Without the folk psychological level of explanation, SFWs do not seem to be in the control of the agent because the network which is activated would seem to do so only as the result of brain activity. SFWs in this case would seem like mere chance and not in the control of the agent. The agent's beliefs, desires, and reasons which conflict and the agent's deliberative efforts occur at the folk psychological level of explanation. Mental causation, which is the causation of actions by beliefs and desires, therefore also occurs at the folk psychological level of explanation. Without mental states and mental causation, the agent's choice will seem to be caused by brain activity (which is not controlled by the agent). Kane argues that agency occurs at the folk psychological level of explanation. (Kane 1998, 147-148)

Explanation at the folk level is the first step in inserting the agent's control into the decision-making process. In the next section, I will go into more detail about the plurality conditions, which must be satisfied in order for an agent to have control over her SFWs.

C. The Plurality Conditions

Kane argues that some of the criticisms involved in the Intelligibility Question are due to what he labels "the problem of plurality." Suppose Jane is deciding where to go on vacation. She has spent days researching various locations and running through different scenarios in each location. She eventually decides that, all things considered at that time, Hawaii is the best spot for her vacation and decides to vacation in Hawaii. In order for this decision to be undetermined, then, with the same prior deliberation, past, circumstances, and laws of nature, Jane could have

decided that, all things considered, Colorado is the best place for a vacation. A theory on free will must explain how this decision to vacation elsewhere is just as rational as the decision to vacation in Hawaii. Problems also arise if the agent's reasons do not incline them towards either choice. Any choice the agent makes seems arbitrary if there is no reason for the agent to choose one over the other. (Kane 1998, 107-109)

Another issue related to the problem of plurality is that indeterminism seems to rob agents of their voluntary control over their actions. If an action is undetermined, then it seems that the action cannot be controlled by anyone. Intuitively, we want to say that, when Jane chooses either Colorado or Hawaii, she as an agent chooses it rather than some uncontrolled quantum events in her brain.

> To have such control over a set of options at a given time is to be able to *bring about* any one of the options (to go more-than-one-way) *at will* or *voluntarily* at the time. That is to say, it is to be able to do *whatever you will* (or most want) to do among a set of options, *whenever you will to do it*, for the reasons you will to do it, and in such manner that neither your doing it nor willing to do it was coerced or compelled (Kane 1998, 111).

Jane must have control when she chooses Hawaii or Colorado given the exact same prior deliberation; it cannot just be a fluke due to a random quantum event in her brain. (Kane 1998, 110-111)

Kane addresses the problem of plurality using his plurality conditions for SFAs. In order to stop the responsibility regress, these SFAs "must be more-than-one-way rational, voluntary, and voluntarily controlled." (Kane 1998, 114) If a decision were only one-way rational and voluntary, then the will would already have been set. Such a decision cannot play the role of an SFA, because, in order for an SFA to function as a basis for the agent's free willing of future decisions, the will cannot be set beforehand. The agent must be able to rationally and voluntarily

choose either way, all past events and deliberation being the same. (Kane 1998, 114) These will-setting decisions are plural, or more-than-one-way, voluntary because the agent sets her own will voluntarily in one way or the other, without being coerced or compelled, and on purpose (e.g. the agent does not perform an action not in line with her will by mistake). She sets her will when she chooses which set of reasons to favor. These decisions are plural rational because the agent possesses a set of reasons for each option. When she chooses an option, she chooses to do the action for the reasons contained in the reason set which corresponds to that option and she makes those reasons the ones she "wanted to act on more than the others *by* choosing for them." (Kane 1998, 135) The agent already possesses the reasons for performing each action prior to conflict in her will. The conflict arises due to the agent's possession of these incommensurable reason sets. (Kane 1998, 114, 135, Kane 2002, 410-412)

Non-SFAs are one-way rational because it would be irrational, based upon the agent's character, past, and motives, for the agent to will otherwise. Kane does not like coffee with cream, so it would be irrational of him to choose to get coffee with cream. Non-SFAs are "one-way" voluntary because the agent would only voluntarily choose one option; the agent's will is already set. Suppose Kane accidently presses the button for coffee with cream because he was not paying proper attention or there was some sort of involuntary spasm in his arm. When Kane pressed the wrong button for his coffee, it was not voluntary, because his will was set such that he wanted the regular coffee and not the coffee with cream. Kane did not have voluntary control because he did not will to push the button for coffee with cream at that time. (Kane 1998, 107-111, 114, Kane 2002, 412)

An agent has control over her SFAs and determinate decisions which trace back to the SFAs by satisfying the plurality conditions when deliberating about an SFA. In the next section,

I will summarize the type of control that Kane requires for the indeterminate SFWs: plural voluntary control.

D. Plural Voluntary Control

For an SFW, no one, not even the agent, is able predict her decision based upon antecedent events in her life and her previous decisions. For SFWs, we lose what Kane refers to as "antecedent determining control," which is the ability to predict the outcome of events based upon antecedent events. For SFWs, an agent instead has plural voluntary control, which is the power to will how one chooses when one chooses, without coercion or compulsion. For non-SFWs, an agent has antecedent determining control from her previous SFWs which formed her characters and motives in order to make it possible to determine the non-SFWs. (Kane 1998, 144, Kane 2002, 420)

Under Kane's account, an agent loses antecedent determining control for some decisions, but she gains ultimate responsibility for these decisions. (Kane 1998, 114, Kane 2002, 420, 431-432) Agents can have control over their choices without having antecedent determining control.

> It does not follow that because you cannot determine or guarantee which of a set of outcomes will occur beforehand, you do not have control over which of them *occurs*, when it occurs...To have plural voluntary control over a set of options means at least to be able to bring about whichever one (of them) you will, when you will to do so, which means in turn that in exercising incompatibilist free will, agents have control over their actions then and there, when they are occurring, even if the actions are not determined by prior events. (Kane 1998, 187)

Kane asserts that assuming that, in order for something to be properly considered a choice or an action, it must be determined by antecedent events begs the question against libertarian theories. Something occurring indeterministically does not necessarily make it a chance occurrence and antecedent determining control is not the only way to control an action. Consider an assassin attempting to assassinate a political figure. Suppose it is

undetermined whether or not he will make a shot that succeeds in killing the political figure due to indeterministic events in his nervous system. First, suppose he fires, hits, and kills the political figure. In this case, we would say that it was the assassin who killed the political figure; the assassin is responsible for the death of the political figure. Now suppose that the assassin fires, but, due to a quantum jump in his arm, he misses. In this case, the assassin would still be responsible for his decision to kill the political figure, even if he did not succeed in killing the political figure. In both cases, the agent chose to shoot the political figure. The assassin is responsible for the willing, even though his ability to accomplish his willing was undetermined. (Kane 2002, 421-423) Kane appeals to Leibniz's dictum that "reasons may incline with necessitating." The assassin had reasons for shooting the political figure, but that does not mean he will necessarily succeed in shooting the political figure. (Kane 1998, 55, 106-107) Consider again the businesswoman. She had two choices: stop to help the mugging victim or continue to her meeting. In this case, she is attempting to successfully carry out each choice. The conflict arises because she cannot both stop and go. When she makes a decision, she made her choice based upon her reasons for each possible choice. It was the result of her own deliberation and she would endorse the choice she made either way as something which she brought about rather than something which was out of her control, even though the choice that she would make was not determined. (Kane 2002, 421-423)

For SFWs, an agent is faced with incommensurable alternatives for which she does not have an all-things-considered preference. She, has for each option, what Kane calls "incommensurable reason sets." (Kane 1998, 166-167) During deliberation, the agent is making

an indeterminate effort to settle this conflict. The reasons causally influence efforts and deliberation, which bring about the choice, which then plays a role in causing the favored action. Since both options correspond to a set of reasons, whichever option she chooses will be rational, even though the choice was not antecedently-determined. (Kane 2002, 416, 420, 422, 426)

It is the loss of antecedent determining control and the gain of plural voluntary control which is needed to introduce indeterminism into this account and to therefore make it a libertarian account of free will. Recall from the previous section agency occurs at the folk psychological level and it is this agency which gives the agent control over this indeterminate deliberation process. The agent, therefore, is in control of this indeterminate decision. (Kane 1998, 146-147)

E. Conclusion: Summary of Kane's Requirements of a Causal Theory

In order to make his account more plausible, Kane requires a theory which can causally link mental states to the decision at the folk psychological level and neural networks to the decision at the neural level for the genuine indeterminate SFWs. He requires a theory which can account for the plurality conditions, rendering the decision of the agent rational rather than random or capricious. The theory must link the agent's mental states with her decision, rendering the decision, whichever one is made, plural rational. It must not force the agent to favor one decision over the other; it must rather allow the agent to make either decision, rendering the decision plural voluntary. He finally needs a theory which can account for plural voluntary control or indeterminate mental causation. The theory must allow for each decision to be made, regardless of the events preceding the decision. In Kane's words, the theory must not require antecedent determining control in order for the decision to be controlled by the agent. In the next chapter I will summarize Woodward's interventionist approach in order to give the

necessary background for my arguments in chapter IV that Woodward's theory can provide Kane's theory with the type of probabilistic causation Kane needs to make his theory more plausible.

Chapter III

Introduction to James Woodward's Interventionist Account of Causation

A. Introduction

In this chapter, I will lay out James Woodward's interventionist account of causation from his book *Making Things Happen*. I will use this information in the next chapter to argue that Kane's libertarian account of free will should use this theory in order to explain the probabilistic causation of SFAs, an agent's plural voluntary control over her decisions, and to account for the plurality conditions.

B. Introduction to Woodward

James Woodward argues that people are interested in causal explanations because they want to be able to manipulate and control events. He further argues that scientific explanation consists of describing how one can hypothetically control or manipulate the outcome of a process. Woodward proposes that good causal explanations provide information on what factors can be manipulated, how they can be manipulated, and under what circumstances they can be manipulated in order to affect the outcome of the process or to affect the probability distribution of the outcome. (Woodward 2003, 9-11)

An intervention is a manipulation of a proposed cause *X* which changes *X* in such a way that a change in the proposed effect *Y* occurs only as a result of the manipulation of *X*. An intervention is used to determine whether there is a causal relationship between *X* and *Y*. Interventions do not need to be performed by humans. There are cases in which a proposed

intervention is not humanly possible to carry out. For example, one cannot intervene to change the value of atmospheric pressure in order to determine if the drop in atmospheric pressure caused bad weather. Counterfactuals made for the purposes of causal explanations must cite what would happen to the value of the proposed effect *if* the value of the proposed cause were intervened upon. In these cases, one utilizes non-experimental evidence in order to determine what would occur if the hypothetical manipulations were carried out. In the next chapter, I will use information about the businesswoman's beliefs, desires, and invariant generalizations about her psychology to propose the effect of hypothetical interventions on her beliefs and desires. (Woodward 2003, 14-15, 35-36)

These explanations rely upon invariant relationships, which are relationships that hold under manipulations of the initial conditions. Invariant generalizations are counterfactuals which give information about the changes on the outcome that would result from a manipulation of the initial conditions. Counterfactuals are used to answer "*what-if-things-had-been-different questions"* or w-questions. (Woodward 2003, 11) A change in the initial conditions is proposed and then the resulting change in the outcome is proposed: "there is some intervention on X such that *if it were possible to intervene to manipulate* X *repeatedly in that way, Y* (or the probability of Y) would change in some reproducible or repeatable way." (Woodward 2003, 71)

These invariant relations are mind-independent. The generalizations hold whether or not humans know these generalizations. Even if humans did not exist, these generalizations would still hold. First, our practice and success in attempting to discover these generalizations would not make sense if these generalizations were mind-dependent. Second, Woodward asserts that it is common practice to attempt to manipulate an outcome by manipulating the initial conditions. If the generalizations on which these tests relied were mind-dependent, then this practice and its

consistent results also would not make any sense. In other words, if these generalizations were mind-dependent, unless each experimenter shared the exact same set of generalizations, then the results of these experiments would not be consistent. (Woodward 2003, 9-11, 71, 118-120)

For Woodward, a causal explanation must contain information about how to, in principle, test the causal claim. He distinguishes between "explanation," which is a human activity concerned with giving information, and "causal relationships," which exist in nature. A causal explanation must latch on to these causal relationships in nature; they must be epistemically accessible. (Woodward 2003, 23)

Consider Newton's second law of motion: "the acceleration of an object is directly proportional to the net force acting on it." (Serway and Faughn 2003, 85) Newton's second law is also an invariant generalization. It holds under the manipulation of the forces acting upon the object in question. The acceleration can be calculated by dividing the sum of the forces acting upon the object by the object's mass:

$$\frac{\text{net force}}{\text{mass}}$$

Suppose a person, Joe, were applying a constant force of 20 N to crate with a mass of 30.6 kg. The acceleration of this object would be 0.654 m/s^2. (Serway and Faughn 2003, 85, 96) Counterfactually, it can be said that if Bob, who is of equal strength to Joe, aided Joe pulling the crate, resulting in a constant force of 40 N now being applied to the crate, then the acceleration of the crate would be 1.31 m/s^2. The proposed intervention is the addition of Bob's force to Joe's force, which would change the cause variable from 20 N to 40N. The invariant generalization being used is Newton's second law. The outcome of the proposed intervention is the value of the crate's acceleration increasing from 0.654 m/s^2 to 1.31 m/s^2; the crate would accelerate faster.

Now consider the counterfactual, "If the barometer reading were to fall, a storm would

occur." (Woodward 2003, 14) If a barometer is operating correctly, then the reading of the barometer falling is correlated with the occurrence of a storm. However, manipulating the reading on the barometer will not alter whether or not a storm will occur. If I am planning a picnic and see the reading of my barometer fall, then manipulating the apparatus so that it has a higher reading will not prevent the storm from occurring. The storm will occur independently of my manipulation of the reading on the barometer. Therefore, the counterfactual under consideration is not a causal explanation because it does not give information on how to manipulate the occurrence of the storm by manipulating the barometer because manipulating the barometer will not change whether or not the storm occurs. (Woodward 2003, 14-15)

C. Interventions

For Woodward, a causal explanation must give hypothetical counterfactual information about how manipulating one event caused a change in another event. When we manipulate the value of a proposed causal variable, we are performing an intervention. Interventions need not be human activities. Woodward asserts that:

> an intervention on some variable *X* with respect to some second variable *Y* is a causal process that changes the value of *X* in an appropriately exogenous way, so that if a change in the value of *Y* occurs, it occurs only in virtue of the change in the value of *X* and not through some other causal route. (Woodward 2003, 94)

In order to establish whether there is a causal relationship between *X* and *Y*, it must be shown that an appropriate manipulation of *X* (one which meets the above criteria) correlates with a change in *Y*. In other words, if a counterfactual asserting that, if *X* is manipulated in a certain way, then *Y* or the probability distribution of *Y* would change is true, then there is evidence for a causal relationship between *X* and *Y.* (Woodward 2003, 105)

In the next section, I will first propose a hypothetical intervention upon the

businesswoman's beliefs and desires in order to establish a causal link between her beliefs and desires and her possible decision, which will show how an agent's choices can be caused by her mental states. I will also illustrate that it is the simultaneous activation of the conflicting mental states which causes the businesswoman's conflict and deliberation. I will second propose a hypothetical intervention upon the agent's self-networks, showing how it is the activation of a particular self-network which causes a certain decision and action and that it is the simultaneous activation of conflicting self-networks which causes the movement away from thermodynamic equilibrium.

D. Invariant Generalizations

Invariant generalizations, unlike the traditional view of scientific laws, can have exceptions. In order for a relationship to count as causal, it must be invariant under at least some background conditions. This means that the generalization will hold when at least some changes or interventions are made under at least some background conditions. A generalization that breaks down all or most of the time when changes are made in all or most background conditions is not invariant and cannot be used to make a causal claim. (Woodward 2003, 147, 239) However, a generalization can still be considered invariant even if it does not hold outside of its domain, outside of a particular spatiotemporal region, or under extreme conditions. So long as it consistently holds under similar background circumstances, it can be considered an invariant generalization. For example, the ideal gas law $PV = nRT$ will not hold under extremely high temperatures; intermolecular forces will interfere with the gas' usual behavior. Invariance is a scalar notion: some generalizations are more invariant than others. Einstein's theory of general relatively is more invariant (holds under a wider range of changes) than generalizations about human behavior. When describing an invariant generalization, the conditions under which it is

invariant must be specified. The generalization must be relativized to certain conditions and to certain systems. The ideal gas law, for example, operates within a system of gases; it will not hold in a liquid system. (Woodward 2003, 239-240, 243-244, 250-251)

Invariant generalizations must be change-relating. Change-relating generalizations give information on how changing the value of a variable or variables will change another value of a variable or variables. These generalizations will answer Woodward's what-if-things-had-been-different questions. Invariant generalizations give us information on how to manipulate a cause in order to manipulate an effect. These relationships are "hypothetically exploitable for manipulation and control." (Woodward 2003, 253, 245-247)

Invariant generalizations can link causes and effects together deterministically or indeterministically. *X* and *Y* can be indeterministically linked when the same change in the value of *X* causes a similar change in the probability distribution of *Y* on multiple occasions. It can also be the case that the same manipulation of *X* under certain circumstances will cause the probability of *Y* to increase rather than definitively causing *Y* to occur. However, this manipulation of *X* may not cause the probability of *Y* to increase under all circumstances and may not cause the probability of *Y* to increase the same amount even in the same circumstances. In nondeterministic situations, there will be an imperfect correlation between *X* and *Y.* (Woodward 2003, 71-72)

In brief, in an indeterministic process, if *C* is causally related to *E*, then, when we intervene upon the probability of *C*, there will be a change in the probability of *E*. It could also be the case that changing the variable of *C* from 0 to 1 or from 1 to 0 can have an effect on the probability of *E*. In both cases, if *C* is causally connected to *E*, then there will be an invariant relationship between *C*'s probability and *E*'s probability. (Woodward 2003, 245)

In the next chapter, I will first argue that the laws of folk psychology are invariant generalizations. I will second argue that the conflicting beliefs and desires of the agent are the input for the conflicting invariant generalizations which are activated in cases of SFWs like the businesswoman's. Third, I will use invariant generalizations about the neural networks within the self-network to illustrate causation at the neural level. The input at this level would be chemical and electrical.

E. Token Causation

I will now discuss how Woodward's account handles instances of token causation. Woodward's interventionist approach accounts for both type and token causation. Consider my above example about the acceleration of crates. The type causal claim would be the formula:

$$\frac{}{}$$

The token causal claim would be: Joe's force of 20N caused the 30.6 kg crate to accelerate at rate of 0.654 m/s^2. Token causal relations focus upon the relations amongst these actual values x (20N and 30.6 kg) and y (0.654 m/s^2). We ask what would happen to the value of y in a counterfactual situation in which all of the other variables in V (any other variables in the system which could have an effect on the value of y) were fixed at their actual values and the value of X was changed from its actual value x_1 to a different value x_n. (Woodward 2003, 74-75)

We must consider two types of counterfactuals in order to establish token causation. The first is of the form "if *c* had not occurred, then *e* would not have occurred", which Woodward calls "not-not" counterfactuals. (Woodward 2003, 211) These not-not counterfactuals hold in both deterministic and non-deterministic contexts. This not-not counterfactual is giving information about what would occur should someone intervene in such a way as to cause the antecedent event not to occur. (Woodward 2003, 210-212)

The second counterfactual we must consider explicates what will happen if the cause *does* occur. In deterministic cases, this second counterfactual is of the form "If *c* were to occur, *e* would occur." (Woodward 2003, 213) This counterfactual follows from a true causal claim that *c* caused *e*. This counterfactual, according to Woodward, asserts a reproducible deterministic relationship between a cause and an effect. All of the circumstances must be similar. This counterfactual can be expanded to: "If, in similar background circumstances, a similar *c* were to occur, then a similar *e* would occur." (Woodward 2003, 213-214)

In indeterministic contexts, we would consider a more complex counterfactual explicating what happens when the cause does occur:

> Intervening to introduce a *c*-like event into a situation in which a *c*-like event would not otherwise occur changes the situation from one in which an *e*-like event would not have occurred to one in which an *e*-like event sometimes occurs in each of a relatively broad range of background circumstances. (Woodward 2003, 215)

In brief, to establish the token causation of *X* to *Y*, we must hold all of the variables which are direct causes of *Y* but are not on the directed path from *X* to *Y*, Z_i, fixed at values that would not cause a change in the actual outcome of *Y*. When testing the route from *X* to *Y*, we want to make sure that the only changes on *Y* would come from *X*. We must then test all possible combinations of other values that *X* can take and determine if changes in these values correlate with changes in *Y*. We would then fix Z_i at a different combination of values and repeat the same test on the variables on the directed path. *X* is an actual cause of *Y* if a change in the value of *X* is correlated with a change in *Y* for at least one combination of values of Z_i and one combination of values on the *X* to *Y* directed path. (Woodward 2003, 83-84, 214-217)

In Section IV, I will argue that SFWs are instances of token causation, since they involve input unique to a particular situation. I will run through the three above counterfactuals in order

to establish mental causation, neural causation, and probabilistic causation of the agent's decision.

F. Levels of Explanation

The manipulationist account does not favor any level of explanation. One can give a manipulationist explanation of an occurrence without appealing to the underlying mechanisms of the occurrence. Consider again the ideal gas law $PV = nRT$. This law can allow us to answer w-questions about how the manipulation of the temperature of a gas will affect its pressure when the volume is held fixed. The ideal gas law does not give us information about why manipulating the temperature of a gas in a fixed container correlates with a change in pressure. However, so long as a generalization allows us to make counterfactuals about how interventions on the cause(s) will alter the outcome of a process, it counts as a causal explanation. So, the ideal gas law is just as explanatory, within its own domain, as its underlying mechanisms: the behavior of molecules as described by Newtonian mechanics. The lower-level explanation of the behavior of gas molecules provides a finer-grained explanation of this behavior and will answer a wider variety of w-questions. Each level of explanation explains in the same way: by making counterfactuals which allow us to answer w-questions. Since we can give explanations about how intervening upon the input entities correlates with a change in the outcome at multiple levels of explanation, then, using Woodward's approach, we can say that each of these input entities at each level causes the change in the result at that level. (Woodward 2003, 221-223)

Recall from Chapter II that, for Kane, each level of explanation is important. Kane argues that agency is established at the folk psychological level, where the agent's beliefs, desires, and other mental states lie. It is at the folk psychological level that the agent feels that she is making an effort to settle her conflict of will. For Kane, the indeterminacy of the decision

arises from the neural level. (Kane 1998, 128-130, 146-147) However, I will diverge a bit from this in order to illustrate that Woodward can give Kane indeterminacy at the folk level as well as the neural level.

Chapter IV

Using Woodward's Approach to Make Kane's Libertarianism More Plausible

A. Introduction

In this chapter I will argue that Woodward's interventionist approach can be used by Kane to make his theory more plausible. I will focus upon Kane's businesswoman on her way to a meeting. In Section B, I will first test the counterfactuals from the token causation section of the last chapter in order to establish causation at the mental level. I will next argue that these interventions show how the businesswoman's decision fulfills the plurality conditions and is plural voluntarily controlled. In Section C I will test the token causation counterfactuals on the neural level. In Section D I will argue that Woodward's account can provide probabilistic causation at both the neural and the folk level of explanation and briefly argue that this may be an improvement on Kane's original theory. In Section E, I will briefly illustrate that Woodward's approach can also establish causation for non-SFWs.

B. Mental Causation

Commonsense folk psychology consists of a set of counterfactual generalizations with intentional rather than quantitative parameters. I will briefly argue that the laws of folk psychology qualify as invariant generalizations under Woodward's theory.[2] Consider the explanation of an action based upon reasons. Reasons consist of beliefs, desires, and other

[2] The paper from which I derived this explanation of folk psychology, Terence Horgan's *Nonreductive Materialism and the Explanatory Relevance of Folk Psychology*, used Woodward's account as a background to argue that the laws of folk psychology are causal laws that can be used to explain mental events non-reductively. I am only adding a bit more detail to Horgan's argument.

mental attitudes. We can use the following generalization to explain a person's action based upon her reasons: "(*S*) (*D*) (*A*) (If *S* wants *D* and *S* believes that doing *A* will bring about *D*, then *ceteris paribus*, *S* will do *A*)." (Horgan 1993, 296) In other words, if Susan wants a snack and Susan believes that going to the kitchen will bring about Susan getting a snack, then, *ceteris paribus*, Susan will go to the kitchen.

Horgan offers other counterfactual generalizations of folk psychology:

> (*x*) (*p*) [If Wants (*p*)*x* & Discovers (*p*)*x*, then *ceteris paribus*, Pleased (*p*)*x*]
> (*x*) (*p*) (*q*) [If wants *p*(*x*) & Believes (If *q* then *p*) *x*, then *ceteris paribus*, either Wants (*q*)*x* or (∃*s*) (Wants (*s*)*x* &Believes (if *q* then –*s*)*x*] (Horgan 1993, 299)

These are change-relating in that they explicate the relationship between either mental states and actions or mental states and other mental states. They allow us to draw conclusions about how a change in the antecedent (the mental state) will lead to a change in the consequent (the action or second mental state), when all other background conditions are the same. The *ceteris paribus* clause functions to isolate the proposed cause and its proposed effect by fixing the background circumstances such that they do not interfere with the causal claim being tested. (Horgan 1993, 298-299) Recall that for Woodward's account, when intervening upon a possible cause in order to determine if there is a causal relationship between *X* and *Y*, one must fix the other variables in the system in such a way that they did not interfere with *X*'s proposed effect on *Y*. (Woodward 2003, 9-15)

An agent's action is usually explained by her reasons or motives for performing that action. Reasons are the psychological attitudes of the agent, such as her beliefs, desires, expectations, and knowledge, which cause the agent to perform a certain action. For Kane, these psychological attitudes can cause actions and choices. For SFWs, the mental processes involved in the psychological attitudes are indeterminate, which makes the causation by these processes

probabilistic. (Kane 1998, 28, Kane 2002, 416, 426)

Woodward's manipulationist theory can be used to causally link the agent's mental states with her decision. As I argued above, the laws of folk psychology are invariant generalizations which govern an agent's mental states and actions. These generalizations, as well as an agent's reasons, beliefs, and desires, are embedded in her self-network. (Kane 2002, 423) SFWs are token instances of causation because they involve input specific to the particular situation: the agent's reasons at that time for making each possible decision. SFWs arise when an agent is confronted with a situation in which two invariant generalizations and the reasons contained in the antecedent of those generalizations conflict. One can also think of an SFW as a situation in which two or more non-SFWs arise simultaneously and come into conflict. Without the conflicting generalizations and reasons, there would not be a deliberation about which action to favor. If the agent, in this non-conflicting case, *did* make an effort to choose, it would be a case of the agent attempting to discover what she already favored. She has an all-things-considered preference for one of the options, she just has to recall which option she already favors. An effort is made, but the outcome is antecedently-determined. The indeterminacy lies in the deliberation between the conflicting options and not between reason(s) and their corresponding actions (Kane 2002, 412).

The reason or set of reasons which correspond to each generalization are the inputs to each generalization which can be intervened upon. Consider the above case of the businesswoman on her way to a meeting. She is in possession of two now conflicting invariant generalizations. The first, the moral generalization, is something like:

> (A) If I believe that action *X* will help someone in danger and if I can do action *X*, then, *ceteris paribus*, I will decide to do action *X*.

The second, the selfish generalization could be:

> (B) If I believe that an action *Y* will further my career and I can do action *Y*, then, *ceteris paribus*, I will decide to do action *Y*.

In this case, the input to the first moral generalization would be: "stopping to help the mugging victim." The input to the selfish generalization would be: "moving on to arrive to my meeting on time." Now that these generalizations are in conflict, each decision has a certain probability of being chosen which depends upon the character of the agent and her past decisions. If they were not in conflict, then this would be a non-SFW. If the businesswoman had not encountered the mugging, then it would be determined by her character and motivations, all other background conditions being held the same: she would attempt to arrive to the meeting on time. Recall from the section on Woodward that, in order to establish token causation, we can use the not-not counterfactual and the counterfactual asserting what will happen in similar conditions. Since Kane wants a person's decision to be caused by her reasons for the decision, I will first consider each possible decision separately. First, consider her belief that stopping to help the victim would protect the victim from danger and the resulting decision to stop. Suppose we intervened (hypothetically) on her belief that stopping would help the victim. Instead, she believes that stopping would not help the person but rather put herself in danger. Plugging this new information in to the above invariant generalization would yield the result that she would not stop to help the person because stopping would actually not help the person. Since we must keep all other background conditions the same, then she must still hold her selfish generalization. This would mean that she experiences no conflict and continues on to her meeting. Now suppose we intervened (hypothetically) on her belief that making it to that meeting would further her career. Again, plugging this new input into the above generalization would yield the result that she would then not be in such a rush to get to the meeting! Again, keeping all other background conditions the same (she does believe that stopping to help the victim would save

that victim from danger), there would no longer be a conflict between her moral and her selfish generalizations, and she would stop and help the victim. We would also note that it was the activation of both reason sets and the possession of both generalizations which caused the deliberation, since, when only one reason set and one generalization was activated, there was not any conflict.

Next I will propose hypothetical interventions using the second counterfactual from the token causation section: "If *c* were to occur, *e* would occur." (Woodward 2003, 213) In the case of the businesswoman, we would again test two counterfactuals:

> (C) If I believe that stopping to help the mugging victim will help someone in danger, than, *ceterus paribus*, I will decide to stop to help the mugging victim.
> (D) If I believe that arriving to my meeting on time will further my career, then, *ceterus paribus*, I will decide to arrive to my meeting on time.

In order to test these counterfactuals, we would first have to intervene (hypothetically) in such a way as to keep the background circumstances from interfering with the effect. First, I will test counterfactual *C*. We would set the background circumstances in such a way that she was not actually on her way to the meeting. In this case, there would be nothing interfering with her moral belief that she should help someone in danger, and she would therefore decide to stop to help the mugging victim. In the case of counterfactual *D*, we would intervene in such a way as to prevent her from encountering the mugging in the alley. In this case, again, there would be nothing interfering with her resolve to arrive to her meeting on time and she would decide to continue on to her meeting. Finally, as when we tested the not-not counterfactual, we would also note that it was the two counterfactuals being activated in conjunction which causes the interference and the deliberation.

Now consider the reason sets and generalizations together. In this case, it is the

background conditions which cause the above generalizations to conflict. Stopping to help the mugging victim would cause the businesswoman to be late for her meeting. First, we can use interventions to show that it is the agent's conflicting sets of beliefs and desires which causes the conflict and deliberation to arise. In this case, the hypothetical intervention would prevent the indeterministic interference itself from arising. This can be done in a variety of ways, some more practical than others. First, we can intervene upon a situation in which the businesswoman is on her way to a meeting and does not encounter a mugging. Prior to the intervention, holding all other background circumstances in such a way that they do not interfere with the agent's belief about the meeting and her desire to further her career, the businesswoman's decision to arrive at her meeting on time would be a non-SFW. We can intervene in such a way as to make her encounter the mugging. In this case, her moral beliefs and desires would interfere with her selfish beliefs and desires; she would now have to deliberate between her two conflicting beliefs and desires. This would now be an SFW. Second, we can intervene on a situation in which she encounters the mugging, but she is not on her way to an important meeting. As above, prior to the intervention, this would be a non-SFW: the businesswoman would non-deliberatively (if she did make some sort of deliberative effort, it would be deterministic) decide to stop to help the mugging victim. We would intervene in such a way as to make it such that she is on her way to an important meeting. Again, her selfish beliefs and desires would interfere with her moral beliefs and desires and she would now have to deliberate indeterministically. In both of these situations, when we intervened and introduced the events which caused the conflicting counterfactuals, we saw the situation changed from one in which she did not have to deliberate (non-SFW) to one in which she did have to deliberate (SFW). Ultimately, whichever one is chosen is caused by the reasons associated with each choice, as illustrated by interventions

performed on them earlier in this section on each set of reasons and their corresponding actions.

Second, we also must consider the counterfactual which Woodward associates with indeterminate token causation in order to establish the indeterminate causation of the agent's decision:

> Intervening to introduce a *c*-like event into a situation in which a *c*-like event would not otherwise occur changes the situation from one in which an *e*-like event would not have occurred to one in which an *e*-like event sometimes occurs in each of a relatively broad range of background circumstances. (Woodward 2003, 215)

The *c*-like event in this test would be the mutual activation of the businesswoman's conflicting beliefs and desires from the second set of counterfactuals, *C* and *D* with their corresponding probabilities of being favored in relation to the probability of the other counterfactual being favored:

$$P(C \bullet D) = 0$$
$$P(C) + P(D) = 1$$
$$P(C) = 0.7$$
$$P(D) = 0.3$$

The *e*-like event would be the resulting conditional probabilities. In this set of equations, the first two equations illustrate the relation between the two set of beliefs and desires. The first equation states that she cannot favor both sets of beliefs and desires. The second equation asserts that she will favor one or the other sets of beliefs and desires. Since the set of beliefs and desires being favored *is* the decision, (Kane 2002, 412) the resulting probabilities would mirror the probabilities of the inputs:

$$P(\text{Decide C} \bullet \text{Decide D}) = 0$$
$$P(\text{Decide C}) + P(\text{Decide D}) = 1$$
$$P(\text{Decide C}) = 0.7$$
$$P(\text{Decide D}) = 0.3$$

In this second set of equations, the first two equations illustrate the relation between the

decisions. The first equation states that she cannot decide to perform both actions at the same time. The second equation asserts that she will decide to perform one of the actions. Both of the choices must be considered because, without the interference of the businesswoman's beliefs and desires about her career advancement, the probability that the woman will decide to stop would be 1, which makes it determined that she will decide to stop. In order to test the above counterfactual, we would have to introduce the *c*-like event in a situation in which it would not otherwise occur. So, we would have to activate the same conflicting set of beliefs and desires, retaining the same probability that each will be favored, in order to test the causal link between the set of probabilities of those beliefs and desires and the resulting probability set. The same probabilities must be retained because we are attempting to explain this particular instance of mental causation. A type causal explanation would not explain this instance because the information it gives is prior to the choice and after the choice and not of the choice itself. In brief, a type causal explanation would tell us that there is a probability of 0.7 that the businesswoman would favor the moral generalization (of which her beliefs and desires at that moment comprise the input) and a probability of 0.3 that she will favor the selfish generalization. No reference is made to the actual beliefs and desires that go into the moral and selfish generalizations. After the choice is actually made, the probabilities of the type causal explanation would adjust to reflect the decision: if the businesswoman favored her moral generalization, this would add more "weight" to that generalization and, in the future, there would be a higher probability that she would favor that generalization again. Due to the fact that probabilities evolve based upon past decisions, it would be impossible to actually perform this intervention. We would have to intervene in such a way as to "reset" the businesswoman's set of probabilities to their original values before she made her decision. We would then have to place

her in a similar situation: she is on her way to an important meeting and she encounters a mugging in an alley. We would then note that, given the same set of probabilities that she will favor each set of reasons, there is the same set of probabilities that she will perform the corresponding actions, since favoring a set of reasons *is* forming the intention to perform a certain action, which then leads to the action itself.

To further illustrate the probabilistic links between the businesswoman's reasons and her action, suppose we intervened on the probability that she will favor the moral generalization in the same background circumstances: she is on her way to an important meeting when she encounters the mugging, the same generalizations will be activated, and the same reasons will serve as the input to those generalizations. Suppose we altered the probability that she will favor her moral reasons from 0.7 to 0.6. Since, according to our background conditions, P(C) + P(D) = 1, then the probability that she will favor her selfish reasons must also change:

P(C • D) = 0
P(C) + P(D) = 1
P(C) = ~~0.7~~ 0.6
P(D) = ~~0.3~~ 0.4

This would correspond with a change in the probabilities that she will favor each decision:

P(Decision C • Decision D) = 0
P(Decision C) + P(Decision D) = 1
P(Decision C) = ~~0.7~~ 0.6
P(Decision D) = ~~0.3~~ 0.4

So, according to Woodward's causal approach, if a manipulation of the input is correlated with a change in the probability distribution of the output, then the input probabilistically causes the output.

In brief, first, we have a causal explanation of each possible action which links the agent's mental states to her decision. Second, it was illustrated that it is the incommensurable

generalizations and reasons held by the agent which caused the conflict and the deliberation. Third, we have a probabilistic causal explanation of the decision that is ultimately made by intervening to cause the same reasons and generalizations to be activated with the same probabilities of being favored. We also established probabilistic causation by showing that there were corresponding changes in the outcome as a result of the interventions upon the probabilities of the initial conditions.

This account meets the plurality conditions. First, since Woodward's account does not stipulate that the option with the highest probability is the one which will be chosen, it does not impinge upon the agent's voluntary willing. She can choose to favor any reason set, no matter what probability is initially assigned to it by her past character and decisions. An option having a higher probability of being chosen does not determine that it will be chosen, the higher probability only indicates that it is more likely that that option will be chosen. So, she can voluntarily will either way, even if she favors the less-probable-for-her selfish reason. Second, the businesswoman had reasons for each possible decision: the moral reason of helping a person in danger as a reason to stop and the selfish reason of furthering her career as a reason to arrive to her meeting on time. The first two interventions which I performed were on those two reasons, which, in keeping with the generalizations of folk psychology, were correlated with a change in her action. The interventions were performed on reasons of the agent and corresponded with changes in her actions. This established a causal connection in both cases between the businesswoman's reasons and her actions. Since either way the businesswoman chose, her action would be caused by her reasons for that action, then her choice is plural rational.

Plural voluntary control is also established. First, since both generalizations and set of

beliefs and desires were simultaneously held by the agent and these generalizations, beliefs, and desires were linked in my explanation above, then the agent chooses for her own reasons either way. The agent in this way controls her decision because her decision is made based upon her own generalizations, beliefs, and desires even though it is not determined by the generalizations, beliefs, and desires. Second, the probability that each decision would be made arises from the agent's character, her environment, and her past decisions. SFWs are not determined by her character, environment, and past decisions, but the probability that she will choose each possible choice also come from the agent herself. Third, the conflict is caused by the beliefs and desires which are held by the agent; it was the agent herself who caused the decision to become indeterminate. Finally, it is the agent herself who is deliberating. Recall from the Kane chapter that, on the folk psychological level of explanation, the agent can feel that she is making an effort to settle a conflict in her will. The effort is something which is done by the agent herself at the folk psychological level of explanation. Since the agent is making an effort to settle her conflict of will, then she is in control of this effort and therefore of the outcome of the deliberation process. In this way, even though the outcome of the deliberation process is not determined by antecedent circumstances and is indeterminate, it is still under the control of the agent and not arbitrary or capricious.

C. Physical Theory

Woodward's causal account will operate on the physical level in the same way it operates on the mental level, except that the conflicting invariant generalizations will be about neural networks and the input to these generalizations will be electrical and chemical. As with the mental level of explanation, without the conflicting generalizations, the outcome would be determined; it is the joint attempt to activate two or more neural networks which causes the

movement away from thermodynamic equilibrium and which renders the outcome (which network will reach activation) indeterministic. The two neural networks are competing to activate. Whichever one is activated is the one which receives sufficient chemical and electrical input.

I will again consider the businesswoman. Not being a neuroscientist, I will use two simple invariant generalizations:

> (E) If Neural Network A receives sufficient electrical and chemical input, then Neural Network A will be activated.
> (F) If Neural Network B receives sufficient electrical and chemical input, then Neural Network B will be activated.

Interventions on the neural activity of the businesswoman's brain would be difficult to do practically, but Woodward does not require that interventions actually occur in order to be explanatory. Suppose we intervene on the electrical and chemical input for Neural Network A in such a way as to prevent it from reaching the network, holding all other background conditions constant, including the electrical and chemical activity flowing into Neural Network B. In this case, Neural Network A would not be active in any way and therefore would not interfere with Neural Network B. That particular region of her brain would not move away from thermodynamic equilibrium and she would not experience any conflict. Neural Network B would therefore be activated without any indeterministic interference. Now suppose we intervene on the electrical and chemical input for Neural Network B in the same way, while allowing input to A. The result would be the same, except it would be Neural Network A which would immediately be activated. These interventions would first show that it is the input into the networks in the businesswoman's self- network which caused the activation of each of these networks. These interventions would also show that the indeterministic conflict only arises when both networks receive sufficient electrical and chemical input. (Woodward 2003, 214-216)

Now consider the counterfactual: "If a similar *c* were to occur (in similar background circumstances), then a similar *e* would occur." Suppose we hypothetically intervene on the businesswoman's neural activity when she is in similar circumstances: she is on her way to a meeting and she encounters a mugging. Suppose we hold all of her neural activity just prior to the activation of the conflicting networks fixed. First, suppose we fixed her neural activity in such a way such that only Neural Network A receives sufficient chemical and electrical input while Network B does not receive any activating chemical or electrical input. Without the competition from Neural Network B, A would activate without interference and the businesswoman would favor the reason network encoded in Network A. Second, suppose we fixed her neural activity in such a way that Neural Network B receives sufficient chemical and electrical input, while Network A does not receive activating chemical or electrical input. Again, without the competition from Neural Network A, B would activate without interference and the businesswoman would favor the reason network encoded in B. In considering this counterfactual and the above not-not counterfactual, we can link the input into the networks with their corresponding action. We also see that it is the simultaneous activation of the networks which leads their interference with one another.

Now consider the two networks together. First, again, consider the counterfactual in the previous passage. First, suppose we held fixed that Neural Network A receives sufficient chemical and electrical input. All by itself, this would mean that Neural Network A would activate. Now suppose we intervened in order to cause Neural Network B to simultaneously receive sufficient chemical and electrical input. Now we would see that neither network activates right away. There would be a competition between the two networks for the chemical and electrical input. If Neural Network A manages to, at the end of the competition, receive the

extra needed chemical and electrical input, then Neural Network A will activate and Neural Network B will not activate. Second, suppose we now held fixed that B receives sufficient chemical and electrical input. Without any other conflicting network receiving sufficient input, Neural Network B would activate without delay or interference. Now suppose we intervened in such a way as to cause A to simultaneously receive sufficient chemical and electrical input. Now we would again see competition between the two networks and only one would be able to activate. This second set of interventions would show that it is the joint activation of the two networks which causes them to interfere and compete with one another, ultimately leading to one network activating and the other not activating.

I will now consider the indeterministic counterfactual:

> Intervening to introduce a *c*-like event into a situation in which a *c*-like event would not otherwise occur changes the situation from one in which an *e*-like event would not have occurred to one in which an *e*-like event sometimes occurs in each of a relatively broad range of background circumstances. (Woodward 2003, 215)

Since this part of my explanation is similar to my explanation of probabilistic causation in the previous section, I will be brief. Consider this set of conditions, in which A stands for "Network A receives sufficient chemical and electrical input" and B stands for "Network B receives sufficient chemical and electrical input":

$P(A \bullet B) = 0$
$P(A) + P(B) = 1$
$P(A) = 0.7$
$P(B) = 0.3$

Since receiving sufficient input will cause the network to activate, then the probability that each network will activate is the same as the probability that it will win the competition for input. An alteration in the probability that each network will win the competition will be mirrored by the

probability that each network will activate.

P(A will activate • B will activate) = 0
P(A will activate) + P(B will activate) = 1
P(A will activate) = 0.7
P(B will activate) = 0.3

It is not the case that both Network A and Network B can simultaneously receive sufficient chemical and electrical input and activate. One of the networks will receive the sufficient input and will activate. The probability that Network A will receive sufficient input and activate is 0.7. The probability that Network B will receive sufficient input and activate is 0.3. As in the mental level explanation, in order to perform this intervention, one would have to "reset" the businesswoman's self-network to before she made her decision. She would have to placed in similar circumstances and then have each network simultaneously activated with the original weights. In this way, we can link the original input probability distribution with the original output probability distribution since the first two sets of interventions has already established a causal link between the networks receiving sufficient input and their corresponding activation. We can also show how altering the probabilities of the input will affect the output. Similarly to my explanation at the mental level, if I intervene on the probability that Network B will win the competition and change it to 0.2, then the probability that B will activate will become 0.2. This will cause a corresponding change in the probability that A will win the competition and activate in order to retain the background conditions of the system. Since the P(A will activate + B will activate) = 1, P(A will activate) will change to 0.8. These final two interventions illustrate that a change on the input (the probabilities that each network will receive sufficient chemical and electrical input) corresponds with a change in the probability distribution of the output (the probabilities that each network will be activated). For Woodward, this illustrates that these input

and outputs are probabilistically-related.

D. Levels of Explanation

I have already illustrated above how we can answer these w-questions using counterfactuals on both of the levels of explanation required by Kane. Woodward's account can thus provide explanations on both the neural and the folk psychological levels for whichever action the agent ultimately decides to favor. Most importantly, it can provide a probabilistic account on both levels, rendering the decision indeterministic on both levels, which means that Kane does not have to give an account of how each reason and choice reduces to neural activity. The context of the situation dictates which level of explanation is needed to explain an occurrence. Not having to give a reductive explanation has the advantage of avoiding eliminativist temptations and giving the agent increased control over her deliberation. First, recall from Chapter 3, Section F, that, for Woodward, both the mental and the neural level of explanation are equally important and the favored level of explanation depends upon the context in which we are giving a causal explanation. (Woodward 2003, 221-223) Using Woodward's account, Kane can avoid eliminativist temptations, which is the temptation to dispense with the higher level of explanation once we can fully explain it in terms of the lower level of explanation. We can give causal explanations in terms of beliefs and desires at the psychological level. As illustrated above, it is possible to use non-experimental data about the businesswoman's mental states and internal invariant generalizations in order to project what would result from a hypothetical manipulation of her mental states. For Woodward, if the hypothetical manipulation of one entity or event can be hypothetically correlated with a change in another entity or event, then there is evidence for a causal relationship between the two entities or events. We can give causal explanations at both levels without one level cancelling

out the other because these hypothetical manipulations can be carried out on both levels. Second, this can also give Kane indeterminacy at the psychological level on its own without relying on indeterminacy propagating up from the neural level. Recall from the summary of Kane that Kane argues that the agent is in control of her deliberation due to her phenomenological experience of her own efforts to make a choice. (Kane 1998, 128-130, 146-147) Allowing indeterminacy to arise on its own at the psychological level therefore seems to give the agent additional control over the indeterminate deliberation by bringing the indeterminacy to the level of agency without depending upon indeterminacy at the neural level.

E. Non-SFAs

Recall from the first chapter that Kane's approach does not require that all decisions be indeterminate; it does not undermine an agent's free will that some of her decisions are determined by her character, past decisions, and surrounding circumstances (Kane 1998, 15). Woodward's approach can account for both SFAs and non-SFAs.

Establishing causation for non-SFAs is more simple than testing for causation for SFAs. Non-SFAs only involve one generalization and one reason or set of reasons; there is no conflict, not movement away from thermodynamic equilibrium, no indeterministic interference, and no deliberation between conflicting options. Non-SFAs would be similar to the above cases in which I tested the not-not counterfactuals, which suppressed the activation of one of the conflicting neural networks and one of the conflicting reason sets. Consider Kane going for a cup of coffee. He already has an all-things-considered preference for regular coffee. (Kane 2002, 412, 417) We would hold all background conditions the same: the antecedent circumstances which lead to him having this preference for regular coffee, the availability of the machine which sells the different coffees, Kane's physical ability to get the coffee, etc. We can appeal to Fodor's

folk psychological law which links beliefs and desires to actions: "(*S*) (*D*) (*A*) (If *S* wants *D* and *S* believes that doing *A* will bring about *D*, then *ceteris paribus*, *S* will do *A*)." (Horgan 1993, 296) If Kane wants regular coffee and Kane believes that going to the coffee machine and pressing the button for regular coffee will bring about Kane getting regular coffee, then, *ceteris paribus*, Kane will go to the coffee machine and press the button for regular coffee. In this case, we can intervene on either Kane's desire for regular coffee or his belief that he can get regular coffee by going to the machine and pressing the button for regular coffee. First, suppose we intervene on Kane's desire for regular coffee. Suppose we change *D* from "regular coffee" to "espresso." Since his belief about regular coffee is on the directed path from his desire for regular coffee to his going to the coffee machine to get regular coffee, we must also change his belief to one about espresso. Suppose Kane believes that he can get espresso at Starbucks (*A* now represents "going to Starbucks"). We would then, according to the invariant generalization, see a change in Kane's action: he would go to Starbucks in order to get an espresso. Second, suppose we intervene on his belief that he can get regular coffee by pressing the button for regular coffee on the coffee machine. Suppose someone plays a prank on Kane and tells him that the machine is broken and he must press the button for a coffee with cream in order to get regular coffee (*A* now represents pressing the button for a coffee with cream). Again, following the generalization, we would see a change in Kane's action: he would press the button for a coffee with cream rather than the button for a regular coffee. In performing these interventions, a link between the mental states and the action is established because the interventions caused a change in the resulting action. Woodward's account can therefore account for both SFAs and non-SFAs in that it can be used in both cases to link the agent's mental states with her resulting choice and action.

Chapter V

Consideration of Other Causal Theories

A. Introduction

In this chapter I will consider alternative probabilistic causal approaches for Kane in order to defend my choice of Woodward's interventionist theory. The first I will consider is Lewis's counterfactual approach. I will argue that, while it can also make Kane's libertarian view intelligible, Woodward's account can do the same in a more simple fashion because Woodward's account is non-reductive and accounts for type and token causation. The second approach I will consider are those collectively labeled "causal process theories," which include Phil Dowe's conserved quantity account, Salmon's mark transmission theory, and transference theories. (Dowe 2007) I will focus on Dowe's conserved quantity approach as a representative of those theories. I will argue that it cannot give Kane the account of agency at the folk psychological level that Woodward can provide.

B. David Lewis's Counterfactual Causal Approach

Both Woodward and Lewis' account rely upon counterfactuals to make causal claims. Lewis' account, however, only accounts for token causation, whereas Woodward's can account for both type and token causation. Suppose we are attempting to discover if *X* causes *Y*. Under Lewis' account, we would need to establish two counterfactuals:

> If *X* occurs, then *Y* will occur.
> If *X* does not occur, then *Y* will not occur.

The second is the not-not counterfactual appealed to in my account of Woodward's theory in the previous chapter. Lewis appeals to possible worlds to establish the truth of his counterfactuals: if the first counterfactual holds in nearby possible worlds in which *X* is true and *Y* is true and *Y* is not false in any of these nearby possible worlds, then the first counterfactual is true and there is a

causal relation between *X* and *Y* (Woodward 2003, 133-136).

When considering the second counterfactual, Lewis requires a small localized miracle in each of the nearby possible worlds which renders it such that *X*, which does occur in this world, does not occur in that world. This is similar to Woodward's interventions in that both prevent *X* from occurring and prevent *X* from occurring in such a way as to render any change in *Y* solely the result of the change in *X*. Both interventions and small miracles can violate the laws of nature. (Woodward 2003, 133-136)

Lewis' account is a reductionist account of causation: he is attempting to reduce the notion of causation to the notion of counterfactual dependence, which is a noncausal notion. Woodward's account is non-reductive: "According to the manipulationist account, given that *C* causes *E*, which counterfactual claims involving *C* and *E* are true will always depend on which other *causal* claims involving other variables besides *C* and *E* are true in the situation under discussion." (Woodward 2003, 136) Woodward's account is nonreductive because he is not attempting to reduce causation to a noncausal notion. This is different than the claim above that his account does not require reductive explanations between levels of explanation; the claim in this section is a claim about Woodward's theory itself. (Woodward 2003, 133-136)

I will now briefly give an account of the mental causation of an SFW using Lewis' theory. Suppose the businesswoman stops and I am now attempting to prove that she stopped because she favored her moral reasons:

> (G) If I believe that stopping to help the mugging victim will help someone in danger and if I can stop to help the mugging victim, then, *ceteris paribus*, I will decide to stop and help the mugging victim.

For Lewis, the causal claim reduces to the non-causal counterfactual G. I can insert a small miracle during her indeterminate deliberation which alters her belief about helping the mugging

victim: “If the businesswoman does not believe that helping the mugging victim is the morally correct thing to do, then the businesswoman will not decide to stop.” This can illustrate that her belief caused her to perform a certain action, which Kane wants. However, we are not yet done. The situation still contains the unreduced causal claim linking her beliefs about getting to her meeting on time. The counterfactual for this set would be:

> (H) If I believe that getting to the meeting on time will further my career and I can get to the meeting on time, then, *ceteris paribus*, I will decide to get to my meeting on time.

Again, we would use a not-not counterfactual to illustrate a causal claim, “If the businesswoman did not believe that getting to the meeting on time will further her career, then she will not decide to get to her meeting on time.” Again, we have linked the businesswoman’s belief with her action.

We must also account for the conflict between the two sets of reasons and the resulting deliberation. We could insert a small miracle such that she does not encounter the mugging at all and therefore G is not active at all. H would still be activated, but there would not be a conflict, so we would see that it was the encounter with the mugging situation that caused the conflict. To be thorough, we would also have to then go back to the original situation and insert a small miracle such that she is not on her way to an important meeting. Now, H would be activated but G would not be activated. There would be no conflict and no deliberation. She would stop to help the mugging victim. Again, we would see that it was the encountering of the mugging when she was on her way to the meeting that led to the conflict and the deliberation.

I will not continue my account further than this consideration of the not-not counterfactual because I believe that the reductions I have already performed can illustrate my point adequately. Lewis can get us an account of mental causation and the conflict, but this

account is slightly more complicated than the account I gave above using Woodward's approach. Using Lewis's approach, starting with the not-not counterfactuals, required at least four small miracles to explain causation due to mental states and the conflict between the two sets of reasons. Using Woodward's approach I was able to show the causation of decisions by mental states and the conflict between the mental states with the first two interventions involving the not-not counterfactuals. Using Woodward's approach, we only had to intervene on each set of reasons in order to show both mental causation and that reason set's contribution to the conflict since, under Woodward's account, we could retain the causal claim made by the other reason set.

Second, as mentioned above, Lewis' account only accounts for token causation. Kane gives an account of an agent's character development which involves evolving probabilities that certain decisions will be made based upon her past SFWs. If the businesswoman favors her moral reasons over her selfish reasons, then the probability that she will favor her moral reasons in the future is increased. (Kane 1998, 176-178) This account is better explained by appealing to Woodward's invariant generalizations than by appealing to a large number of token events stored in her self-network. An explanation of this phenomenon, under Woodward's approach, can be that, when the businesswoman favors her moral reasons, she is favoring her moral generalization and thus increasing the probability that she will favor this particular moral generalization in the future. The self-network is therefore only storing these generalizations rather than each instance of moral dilemma encountered by the agent. These generalizations are strengthened or weakened depending upon how often the agent favors the reasons which comprise their input.

There are a few other differences between Woodward and Lewis' theories, but, since this project is not a defense of Woodward's theory, I will not focus on the rest. These were the reasons I ultimately favored Woodward's account over Lewis' account: Woodward's account can

give Kane what he needs more simply than Lewis' account.

C. Causal Process Theories

I will now consider Phil Dowe's conserved quantity approach as a representative of causal process theories. Dowe argues that a process is causal if it possesses and transfers a conserved quantity. He argues for two main theses:

> "CQ1. A causal process is a world line of an object that possesses a conserved quantity.
> CQ2. A causal interaction is an intersection of world lines that involves exchange of a conserved quantity." (Dowe, Physical Causation 2000, 90)

The world line of an object is "the collection of points on a space time (Minkowski) diagram that represents the history of an object." (Dowe 2000, 90) Conserved quantities each have their own conservation law. This means that whether or not a quantity is conserved is determined by current science. Conserved quantities are quantities such as momentum, charge, and energy. Two processes intersect when they occupy the same space-time location. An exchange occurs at the intersection when one input conserved quantity and one output conserved quantity is altered. The cause and the effect are linked by the exchange of the conserved quantity. (Dowe, Physical Causation 2000, 90-92, 171)

To illustrate, consider the transmutation reaction:

$$He + \quad N + Q \rightarrow \quad O + \quad H$$

An alpha particle (He) collides with a nitrogen atom (N). Both the alpha particle and the nitrogen atom possess the conserved quantity of charge. Q is the extra energy needed to make the reaction occur. One unit of charge is exchanged from the alpha particle to the nitrogen atom, resulting in an oxygen atom (O) and a proton (H). The lines on the chart represent the

worldliness of each of the objects. The intersection shows where the worldlines of the two objects meet and exchange a conserved quantity (in this case, charge). (Dowe 2000, 92-93)

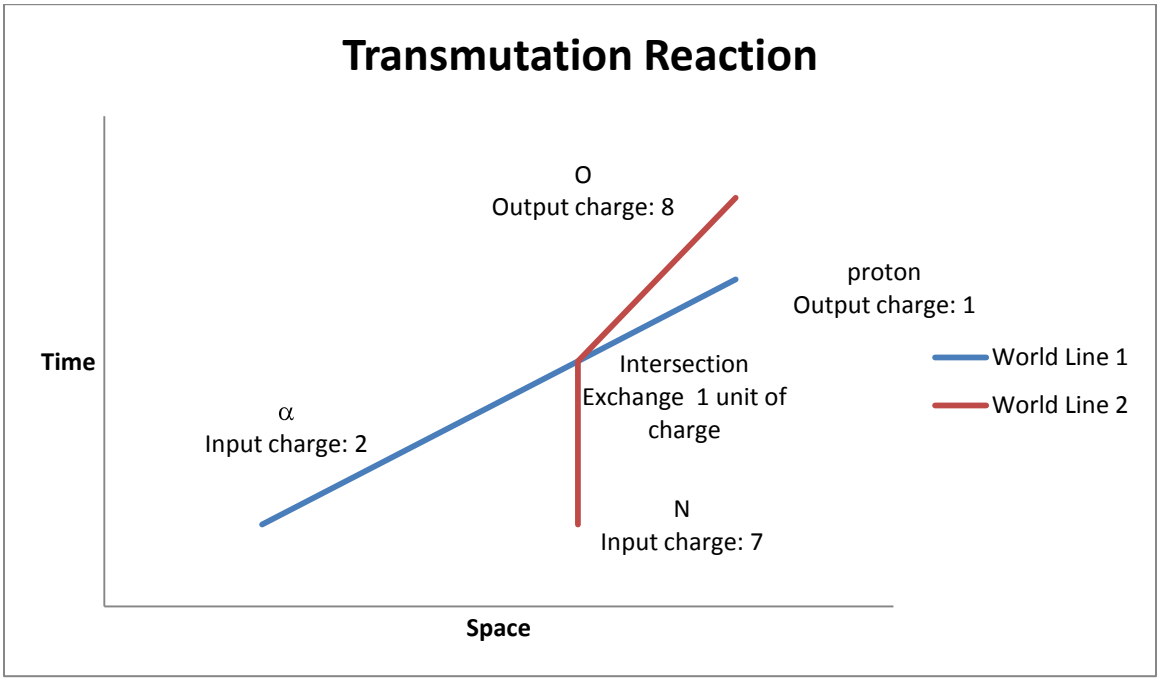

Figure 1: Transmutation Reaction

This approach works for Kane at the neural level. We can refer to the propagation of charge and the chemical activity within and between the neurons. However, there is no conserved quantity which is transferred at the folk psychological level; the laws of folk psychology are not conservation laws. Consider my above example: no quantity is transferred between the businesswoman's desire to help the victim and her stopping to help the victim. As Horgan asserts, "intentional psychological properties are *abstract* properties." (Horgan 1993, 312) Using theories such as Dowe's, we would have to give reductive accounts of mental phenomena, since there is not a conserved quantity which is possessed and transmitted by the mental states. Mental processes would have to be reduced to entities which can possess and transmit conserved quantities, such as neurons and neural networks. The indeterminacy at the neural level would have to propagate up to the psychological level; there would not be

indeterminacy at the psychological level on its own. Kane would thus lose the ability to speak of the agent making an indeterminate effort to make the decision, since he speaks of this effort as occurring on the folk level. The only indeterminism would be at the level of the neurons. The decision would then be the result of indeterministic neural activity rather than the indeterminate efforts of the agent.

Chapter VI

Conclusion

I have argued in favor of using Woodward's interventionist approach in order to help make Kane's theory intelligible. First, Kane's theory requires an account of probabilistic causation which can work on both the neural and the folk psychological level. Woodward's account can provide accounts of probabilistic causation at both levels non-reductively because there are invariant generalizations on both levels and because we can perform interventions on both levels. I ran through interventions on sets of probabilities on both the neural and the psychological level which illustrated the probabilistic link between the probabilities of input (reasons and generalizations or input into neural networks) and the probabilities of the output (an option being favored or a network being activated). Second, Kane's theory requires a causal account of SFWs which fulfills the plurality conditions. Woodward's approach fulfilled the plurality conditions first by allowing us to link the agent's reasons with her decision, thus making them plural rational. Second, Woodward's approach allowed the agents to choose either option, regardless of the probability of that option being favored by the agent, thus making it plural voluntary. Third, Kane needed an account which would make the SFWs plural voluntarily controlled. Woodward's account provided plural voluntary control in a variety of ways: by linking the agent's reasons with her decision, by linking the probability (established by the agent)

that she will favor a certain generalization and reasons to the probability that she will perform corresponding action, showing that the conflict was caused by conflicting reasons and generalizations held by the agent, and by giving the agent control over the indeterminate deliberation process at the folk level of explanation. Finally, Kane's theory requires a theory which can account not only for his non-deterministic SFAs but also for his deterministic non-SFAs. Woodward's account could also provide a causal explanation of non-SFAs because we can perform interventions on the reasons for the determined choice and causally link those reasons to the decision. I considered two alternative causal explanations: Lewis' counterfactual approach and Dowe's conserved quantity approach. I concluded that Lewis' counterfactual approach provided a more complicated causal account for Kane than Woodward's interventionist approach. I concluded that Dowe's conserved quantity approach could not give Kane agency at the folk psychological level. Woodward's approach, in allowing for a non-reductive account of probabilistic causation at the folk level of explanation, seems to give the agent greater control over her indeterminate deliberation.

I am not arguing that bringing in Woodward's causal theory will make Kane's theory the *most* plausible theory of free will; arguing for the most plausible theory of free will is outside the scope of this paper. My conclusion narrower: Woodward can help Kane to better address some of the accusations made against libertarianism. Woodward's approach can causally link the agent's reasons with her decision and help to show how the agent has control over her decision, thus helping Kane to address the Intelligibility Question. In conclusion, Woodward's interventionist approach can be used by Kane to make his libertarian free will approach intelligible by accounting for probabilistic causation and by giving the agent control over her indeterminate decisions.

REFERENCES

Dowe, Phil. "Causal Processes." *Stanford Encyclopedia of Philosophy.* September 10, 2007. http://plato.stanford.edu/entries/causation-process/ (accessed April 20, 2011).

—. *Physical Causation.* New York: Cambridge University Press, 2000.

Horgan, Terence. "Nonreductive Materialism and the Explanatory Autonomy of Pyschology." In *Naturalism: A Critical Appraisal*, by Stephen J. Wagner and Richard Warner, 295-320. Notre Dame: University of Notre Dame Press, 1993.

Kane, Robert. *A Contemporary Introduction to Free Will.* New York: Oxford University Press, 2005.

Kane, Robert. "Some Neglected Pathways in the Free Will Labyrinth." In *The Oxford Handbook of Free Will*, by Robert Kane, 406-437. New York: Oxford University Press, 2002.

—. *The Significance of Free Will.* New York: Oxford University Press, 1998.

Serway, Raymond A., and Jerry S. Faughn. *College Physics.* Pacific Grove, CA: Brooks/Cole-Thomson Learning, 2003.

Woodward, James. *Making Things Happen.* New York: Oxford University Press, Inc., 2003.